Malone et al. *J. Mol. Biol* 253:618-632 (1995)

J. Mol. Biol. (1995) 253, 618–632

JMB

Structure-guided Analysis Reveals Nine Sequence Motifs Conserved among DNA Amino-methyltransferases, and Suggests a Catalytic Mechanism for these Enzymes

Thomas Malone[1], Robert M. Blumenthal[2]* and Xiaodong Cheng[1]*

[1]*W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory Cold Spring Harbor, NY 11724, USA*

[2]*Department of Microbiology Medical College of Ohio Toledo, OH 43699-0008 USA*

**Corresponding authors*

Previous X-ray crystallographic studies have revealed that the catalytic domain of a DNA methyltransferase (Mtase) generating C5-methylcytosine bears a striking structural similarity to that of a Mtase generating N6-methyladenine. Guided by this common structure, we performed a multiple sequence alignment of 42 amino-Mtases (N6-adenine and N4-cytosine). This comparison revealed nine conserved motifs, corresponding to the motifs I to VIII and X previously defined in C5-cytosine Mtases. The amino and C5-cytosine Mtases thus appear to be more closely related than has been appreciated. The amino Mtases could be divided into three groups, based on the sequential order of motifs, and this variation in order may explain why only two motifs were previously recognized in the amino Mtases. The Mtases grouped in this way show several other group-specific properties, including differences in amino acid sequence, molecular mass and DNA sequence specificity. Surprisingly, the N4-cytosine and N6-adenine Mtases do not form separate groups. These results have implications for the catalytic mechanisms, evolution and diversification of this family of enzymes. Furthermore, a comparative analysis of the *S*-adenosyl-L-methionine and adenine/cytosine binding pockets suggests that, structurally and functionally, they are remarkably similar to one another.

Keywords: amino acid sequence motif; catalytic mechanism; DNA methyltransferase; *S*-adenosyl-L-methionine; structural similarity

Introduction

DNA Mtases transfer methyl groups from *S*-adenosyl-L-methionine (AdoMet) to specific positions on bases in double-stranded DNA. The DNA Mtases fall into two major classes, defined by the position methylated. The members of one class methylate a pyrimidine ring carbon yielding C5-methylcytosine (5mC; e.g. *Hha*I Mtase, M.*Hha*I). Members of the second class methylate exocyclic amino nitrogens, forming either N6-methyladenine (N6mA; e.g. M.*Taq*I) or N4-methylcytosine (N4mC; e.g. M.*Pvu*II). Mtases of the two classes were expected to be substantially different from one another, based on the fact that their targets of methyl transfer are very different. This substrate difference can be illustrated by the respective σ-charge densities for the methyl-replaceable hydrogen atoms, which are +0.22*e* on the exocyclic amino groups of both adenine and cytosine, and −0.03*e* on the 5-carbon of cytosine (Renugopalakrishnan *et al.*, 1971). Do Mtases from the two classes, in fact, differ substantially from one another?

Analysis of gene sequences has suggested that the two Mtase classes are quite different. All bacterial 5mC Mtases, and a *Chlorella* virus 5mC Mtase, contain a set of ten conserved blocks of amino acid residues (I through X; Posfai *et al.*, 1989; Cheng *et al.*, 1993a; Kumar *et al.*, 1994; Lauster *et al.*, 1989; Som *et al.*, 1987). These conserved motifs have the same linear order, which simplifies their identification in primary sequences. These ten con-

Abbreviations used: AdoMet, *S*-adenosyl-L-methionine; AdoHcy, *S*-adenosyl-L-homocysteine; Mtase, methyltransferase; 5mC, C5-methylcytosine; N4mC, N4-methylcytosine; N6mA, N6-methyladenine; amino Mtase, Mtase generating N4mC or N6mA; COMtase, catechol *O*-methyltransferase; CM, conserved motif; vdw, van der Waals.

0022-2836/95/440618-15 $12.00/0

served motifs are even present in the carboxy-terminal ~500 amino acid residues of the mouse, human, and *Arabidopsis* CpG 5mC Mtases (Bestor *et al.*, 1988; Scheidt *et al.*, 1991; Finnegan & Dennis, 1993; Guenthner *et al.*, 1992). In contrast, linear alignment of the amino acid sequences of the amino Mtases has not revealed such conservation (see below).

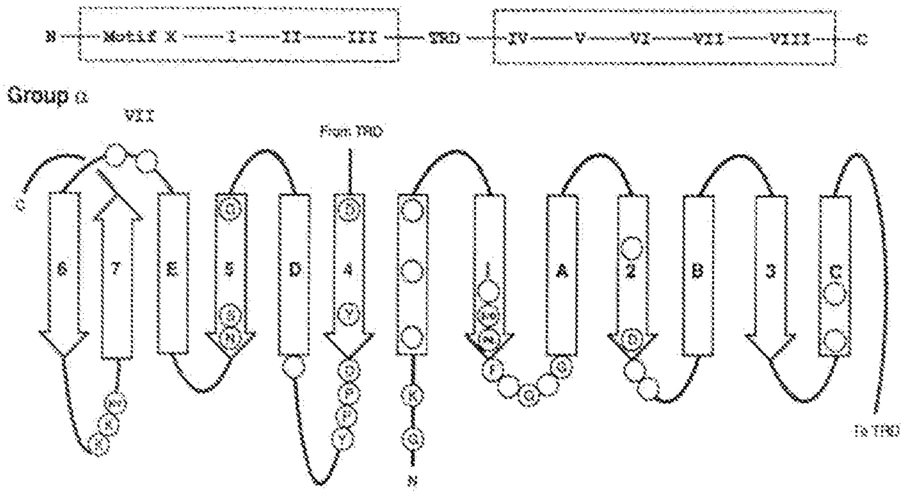
Before any DNA Mtases had been characterized structurally, two motifs of 5mC Mtases were assigned functional roles. Motif I (the core of which is almost always a Gly-rich sequence, such as Ala19-Gly-Leu-Gly-Gly in M.*Hha*I) was presumed to be part of the AdoMet binding site. This assignment was based on the presence of this Gly-rich sequence in a wide variety of AdoMet-dependent Mtases in addition to the 5mC Mtases, including N6mA and N4mC DNA Mtases, and RNA, protein, and small molecule Mtases (Klimasauskas *et al.*, 1989; Ingrosso *et al.*, 1989; Smith *et al.*, 1990; Wilson & Murray, 1991; Kagan & Clarke, 1994). The other motif to which a role could be assigned was motif IV, which contains an invariant dipeptide (Pro-Cys). The Cys in this motif is the active site nucleophile, and forms a transient covalent bond to the 6-carbon of the methylatable cytosine (Santi *et al.*, 1983, 1984; Wu & Santi, 1987; Chen *et al.*, 1991; Friedman & Ansari, 1992; Smith *et al.*, 1992; Wyszynski *et al.*, 1992; Hanck *et al.*, 1993; Mi & Roberts, 1993; Chen *et al.*, 1993). Most amino Mtases lack a Pro-Cys dipeptide.

Structural analysis, however, has found striking similarity between DNA Mtases of the two classes. Information on the structures of DNA Mtases first came from studies of M.*Hha*I and M.*Taq*I, which, like most Mtases from type II restriction-modification systems, are active as monomeric enzymes. These studies have provided insights into Mtase domain organization and its relationship to the conserved sequence motifs (Cheng *et al.*, 1993a; Labahn *et al.*, 1994). The structure of an M.*Hha*I-DNA complex provided further insight into the functions of several conserved amino acids implicated in DNA sequence specificity, catalysis and AdoMet binding (Cheng *et al.*, 1993b; Klimasauskas *et al.*, 1994). M.*Hha*I is folded into two broad domains: a catalytic domain that contains the active site and AdoMet-binding regions, and a DNA-recognition region. The structure of M.*Taq*I complexed with AdoMet is also bilobal (Labahn *et al.*, 1994). This bilobal structure may be a general property of DNA Mtases, as it has also been seen following limited proteolysis of M.*Eco*RI (Reich *et al.*, 1991) and M.*Pvu*II (C. M. Adams & R. M. Blumenthal, unpublished results). In contrast, a single-domain structure has been determined for catechol O-methyltransferase (COMtase: Vidgren *et al.*, 1994). Catechol, like cytosine, is a six-membered ring; this small molecule can readily diffuse into the active site of COMtase for methyltransfer from AdoMet.

The structural comparison of three AdoMet-dependent Mtases reveals that the catalytic domains of the bilobal proteins M.*Hha*I and M.*Taq*I, and the entire single domain of COMtase, all exhibit very similar three-dimensional folding (Schluckebier *et al.*, 1995). The recently published structure of the 5mC Mtase M.*Hae*III (Reinisch *et al.*, 1995) is also consistent with this folding pattern. This similarity includes the positions of amino acid side-chains involved in either AdoMet binding or catalysis. In other words, many of the conserved motifs in the catalytic domain of M.*Hha*I have structural homologs in the other two Mtases (O'Gara *et al.*, 1995). This suggests that many (if not all) AdoMet-dependent Mtases may share a common catalytic domain structure. If so, this not only allows structural predictions for other AdoMet-dependent Mtases, but also provides a framework for attempts to compare their sequences. Guided by this common catalytic domain structure, we performed a multiple sequence alignment of 33 N6mA and 9 N4mC Mtases. Our results reveal that the N4mC and N6mA Mtases are more closely related to one another and to the 5mC Mtases than was expected.

This work confirms that the amino Mtases belong to three groups distinguished by differences in the

Figure 1. Sequence alignment of 33 N6mA DNA Mtases and 9 N4mC DNA Mtases. A, Group α. B, Group β. C, Group γ. Motifs (I to X) are labeled using the nomenclature of Posfai *et al.* (1989), and sequences are grouped (α to γ) using the nomenclature of Wilson (1992). Conserved amino acids are grouped as (E, D, Q, N), (V, L, I, M), (F, Y, W), (C, P, A), (K, R) and (S, T), using standard one-letter abbreviations. Invariant amino acids within a group are shown as white letters against a black background, conserved hydrophobic positions are indicated by bold letters on a shaded background, and conserved polar or charged positions by bold letters within a box. Lesser degrees of conservation are shown, in decreasing order, by bold and uppercase letters, while non-conserved positions are shown as lowercase letters. A (–) indicates a deletion relative to other sequences. Each of the three groups of Mtases is preceded by a theoretical topological drawing. Rectangles (lettered) indicate helices, and arrows (numbered) depict strands. Conserved amino acids from motifs (I to X) are circled and their positions are inferred from the structural comparison of M.*Hha*I and M.*Taq*I (Schluckebier *et al.*, 1995). In addition, the secondary structures of M.*Taq*I, shown in group γ, are indicated by cylinders (helices) and arrows (strands) drawn directly above the amino acids forming them. The amino acid sequences of M.*Hha*I (a 5mC Mtase) and a small-molecule COMtase are also provided in group γ for comparison. (*) Motif X could not be identified for M.*Pse*R7I (N6mA in group γ) using sequence P05103, but was readily found in the sequence used by Wilson (1992), who reported an alternative start position that increases the size of this Mtase from 531 amino acid residues to 574. (**) The Mtases *Fok*I and *Sts*I are each double-size Mtases with two active halves (Sugisaki *et al.*, 1989; Kita *et al.*, 1992), and each half was analyzed independently.



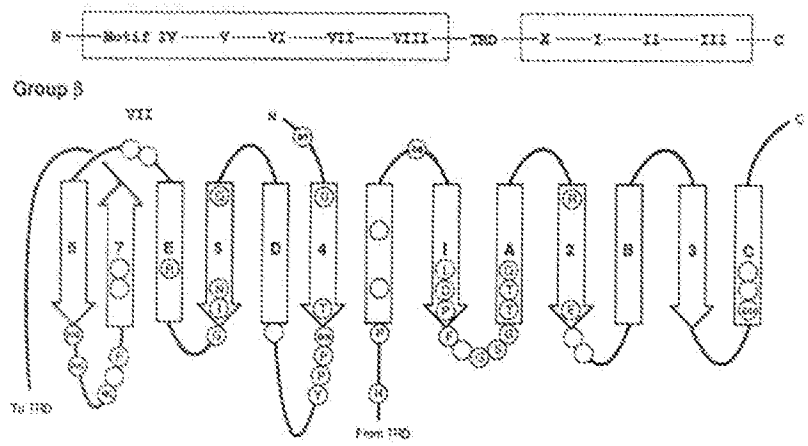
VIII VI V IV X I II III

Protein		Motif X		Motif I		Motif II		Motif III
CviAII (N6)	5	gYiSeilkLkdwIfeeIsK	32	FaDLGsGsciMthEa	54	IsNDLEtypYvDmgL	87	EdckDMVipg
CviBI	6	KWsggkddLkrPedyIPs	29	FIEPgsGAatffhV	51	VlsNvhveLVeLYraI	72	qaiyDPMksh
DpnII 1	16	KWcggkrqLLpvIreIPK	39	YFEPfvGGgaLffDL	59	VINDINseLInnYqqI	79	qeliKILkvh
Dam [Ecoli]	9	KWaggkypLLddIkrhLPK	31	LVEPfvGAGsVflNt	51	IlatDNsdLIsLYniV	88	MrtEMVqas
Dam [T4]	6	aYtgskqsLLpeIksmFPK	28	FVDLfcGGlsVslNV	47	LaNDIQepLIeMYkrL	65	KswIVLkvi
EcoRV	11	KsqWisknLVpcIkrlVPK	35	WVEPfvGtGvVafNV	55	LlcdINphLIsfYnaL	74	ditgDLVkdf
FokI-a	3	RFigsktnLLdnIqeviee	29	FMDLfsGtGIVgeNF	50	LsNDislyfsYIiLkak	89	Hene?elsh
FokI-b	370	nYvgsknkLLnqIvplFPd	393	FVDIfsGGfnVgiNV	413	LaTDINtyVVeVLdtM	427	Kahi?rrise
MboI A	6	KWaggknsLLdeIqkrLPd	33	LVEPfvGGaVslwa	58	VINDINsdLInvYqvI	81	IqyiE?lqsh
NlaIII	2	nYigsklkLsnwIeteIsn	31	F?DLfcGtGIVgrkP	53	IaNDIEyysYvLnrnY	88	Lppr?diyq
StsI-a	2	RYlgsktkLLpeIkkmVdK	29	FLDLfsGtnvVanyF	50	YsNDIlffsYvnakat	83	Iqnl?Vndet
StsI-b	368	nYigsykLLnqIlplFPK	391	FVDIfsGGAnVgiNV	411	IfNDINtrInqMYryF	435	Yqqi?ekide
NvaI (N4)	35	elthryhiYFasMipqLAK	69	LYDPFmGSGTSlvEG	118	aiSDLRysIreMtilY	141	snlpDFdRid

Protein		Motif IV		Motif V		Motif VI		Motif VII
CviAII (N6)	202	tDHeFVDPPYnsrqyqaN	220	YfvlNQIltpkeigng	238	GlpeYk--KSP	265	SarlfviSY
CviBI	172	sNFVFIDPPYdsvftdyg	190	FgksEhVrLaNffktc	210	kclmVig-eTDF	224	DqyihtKY
DpnII 1	188	GDFVYFDPPYiplsetsa	214	FsfsDQVrLrDafkrL	232	tGayVm--lSNS	248	Elykdf?I
Dam [Ecoli]	175	ASFVYCDPPYapisataN	201	FtleQQahLaEisegL	219	rhipVl--lSN?	229	Dtmltr?N
Dam [T4]	165	GDFVYFDPPYlitvadyN	185	WsedEEkdLlNlldsL	203	rGikFg--qSNV	219	Entlli?N
EcoRV	187	GDFVYCDPPYigrhvdyf	207	WgerDErlLfEclsaL	222	lnatFi--tSTW	238	Eskyvr?I
FokI-a	212	GDIIYIDPPYnqrqyisN	230	YhllETIaLyDypsiY	249	GlrpYveskSIY	278	Nfrhilv SY
FokI-b	542	nDLVYLDPPYlitgsyN	568	WnrlQEIkLlDlldhL	586	kGvyFs--lSNV	602	Eneill?N
MboI A	192	FcFVYLDPPYrpisdtaS	218	FddnEQkrLaNfckkI	236	lGhyFl--lSNs	249	Encsed?F
NlaIII	205	GDIIYIDPPYnarqygaN	223	YhllNSIaLyDdftpk	242	GlreYs--rSKY	269	DfqyiflSY
StsI-a	210	ADIVYIDPPYnarqyasN	228	YhllENVarnEhptlk	247	kifdWknlkSDY	276	NsthillSY
StsI-b	540	nDFIYLDPPYlitgsyN	566	WnntEEkkLlNfmdyL	584	hGirFa--lSNV	600	Tnqll?N
NvaI (N4)	267	fDIVVTSPPYgdsKttva	298	Lkid?EtkinQldnvN	315	GGktdk----NI	369	TrvwkpeSY

Protein		Motif VIII	MW	Target	Swissprot Accession Number
CviAII (N6)	299	rnhKRFkaqi	326 aa	CATG	P31118
CviBI	234	ekkyRFklhs	260 aa	GANTC	Q01511
DpnII 1	256	----hYveat	284 aa	GATC	P04043
Dam [Ecoli]	243	hvvkvrrsis	278 aa	GATC	P00475
Dam [T4]	234	hlnKKYvfni	259 aa	GATC	P04392
EcoRV	253	tkehfYhvga	298 aa	GATATC	P04393
FokI-a	315	mpyKKYkakh	**	GATCC	
FokI-b	617	hlqhsYense	647 aa	GGATG	P14871
MboI A	257	-fdelYqdfk	294 aa	GATC	P34720
NlaIII	303	teyKKFkadk	334 aa	CATG	P24582
StsI-a	313	ipyKKYqskn	**	GAGTC	
StsI-b	617	yldnmYnnss	653 aa	GGATG	P29347
NvaI (N4)	400	elfKKYnvhh	454 aa	CCWGG	P14244

Figure 1A

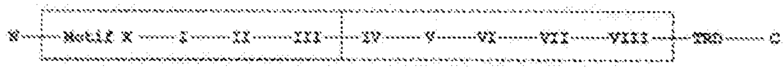


Protein		Motif IV	Motif V	Motif VI	Motif VII
DpnII 2(N6)	41		82		128
EcoI	82		128		174
HinfI	28		71		117
NgoI	21		49		96
PpoI	60		96		145
MboI C	32		73		119
MboII	23		49		95
RsrI	58		81		121
BbaII	30		67		116
BamHI (N4)	137		166		206
BamHII	8		38		66
CfrBI	44		75		115
CfrBI	111		157		191
HchII	82		66		103
PvuII	46		72		112
SmaI	43		73		114
XcyI	44		74		115

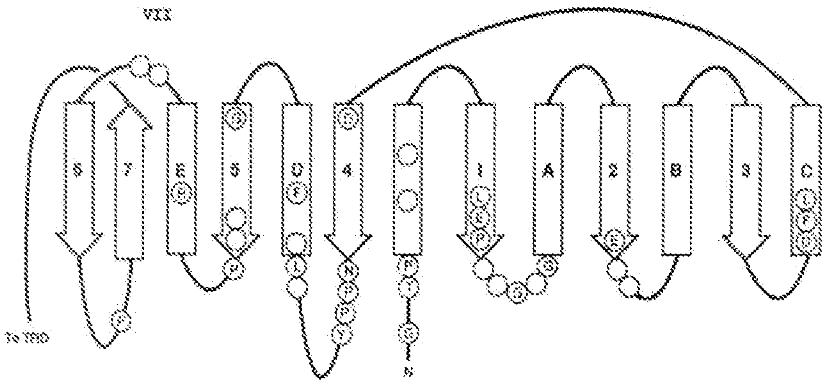
Protein		Motif VIII	Motif X	Motif I
DpnII 2(N6)	137		200	
EcoI	204		182	
HinfI	126		193	
HpaI	105		173	
KpnI	154		218	
MboI C	128		190	
MboII	104		193	
RsrI	143		223	
XbaII	120		169	
BamHI (N4)	220		299	
BamHII	75		182	
CfrBI	140		226	
CfrBI	200		305	
HchII	122		213	
PvuII	122		246	
SmaI	139		227	
XcyI	140		226	

Protein		Motif II	Motif III	MW	Target	Swissprot Accession Number
DpnII 2(N6)	242	259		368 aa	GATC	P08758
EcoI	364	389		531 aa	GAATTC	P16827
HinfI	235	252		359 aa	GANTC	P20590
HpaI	215	232		314 aa	GTTAAC	P29538
KpnI	366	389		417 aa	GGTACC	P25238
MboI C	233	250		373 aa	GATC	P14721
MboII	236	247		360 aa	GAAGA	P23192
RsrI	265	296		319 aa	GAATTC	P14751
XbaII	211	221		228 aa	GATCT	P00673
BamHI (N4)	341	353		423 aa	GGATCC	P23941
BamHII	226	238		365 aa	GGATCC	P18881
CfrBI	268	282		300 aa	CCAGG	P14263
CfrBI	347	368		376 aa	CCAGCC	Q04885
HchII	255	269		355 aa	CTAG	P29568
PvuII	338	313		336 aa	CAGCTG	P14264
SmaI	369	383		293 aa	CCCGGG	P14230
XcyI	368	382		300 aa	CCCGGG	P30776

Figure 1B



Group γ



Protein		Motif X	Motif I	Motif II	Motif III
		VIII VI V IV	X I	II	III
AccI	(N6)	30 AcFFTRfpIayaMakWI	55 VLEPAfGIGvFsrAI	77 niKAfEND	93 eYfdIEsnV
BanIII		16 GshFTPtdIpdIIakrL	49 cLDPscGIgsLLlsL	69 fnnidNIE	87 EsssIEria
BsuBI		26 GsFFTPssIsiFMacLV	53 VLDACaGIgsLtsAP	78 dlhiLEID	98 aLfkDYIeI
CviBIII		26 GlFFTPktVrekLfgPt	53 ILEPscGtgsIIsec	75 siKGVELD	100 IYnsIFLLW
Eco57I		14 GvYYTPgsLadYVtkWV	39 ILEPscGsgvFLgAI	65 cfelFDtE	90 EteqIFLVW
HincII		10 GsFFTPtiDykYMigDk	34 ILEPssGsgvFLdsL	55 nltsYEID	71 VInsIFIts
PaeR7I	^	-10 GslFTrseVvdFIlGLa	32 LLEPsfGsgvFLlPI	83 KreGLdaR	100 LsqgIFLLa
PstI		35 GsFNsssVselMacLF	61 ILDACaGvgsLtsAP	110 KlKAMEtE	124 IIesIFIga
TaqI		19 GrVeTPpeVvdFMvsLa	43 VLEPAcAhGpFLrAP	66 RfvGVEID	85 gIlaIFLLW
TthHB8I		17 GrVeTPpqLvrFMvgLa	41 VLEPAcAhGpFLrAF	64 RfvGVEID	83 gVvaIFLLW
VspI		116 GvFFTPksIsadFfdYL	141 FcDPccGtgsFLiea	181 devALEIt	195 EskrIFLsa
EcoRI		50 PrvenFFKYFavnFdxL	79 nkEGFsssesaknGF	104 KIvfdDIS	123 sesiDLAkk
COMtase		41 nvgdakgqIMdaVIreY	62 VLELGayuGysavrM	88 RlltMENN	117 gaSqDLIpq
HhaI	(C5)	298 GNSVvinVlqyIaynIg	14 FIDLFaGlGgFrlAL	35 cvysnEND	56 kpeGDItqv

Protein		Motif IV	Motif V	Motif VI	Motif VII
AccI	(N6)	117 FDgII-NPPYiKfhdydn	184 NLYtlFlKsI-kqLSqnGrcaYIIPseFL		190 VKcyIikakrIrtII
BanIII		132 VDlIIAnPPYvKtqvlga	166 DLYhefLva-MtlqLKpGGlIgVItsnkYL		202 TKqfLasnydIleIK
BsuBI		137 FThaIlNPPYkKiksnsK	168 NLYsafValtV-dlMcdGGeIvFIIPrsFc		204 FKqhIinktaIkhMn
CviBIII		113 FDIIVGNPPYvvrpegyK	141 NLYvelsyKcItshLK-gGiLsFIIPstIg		178 IKkiIitidiLsfeI
Eco57I		110 FDqaLGNPPYiRyqfleR	149 NsWvpIilssL-alLKqGGrIqMVIPssIs		185 LKsyLqhvcskivII
HincII		85 YDnIIGNPPYvRwknlse	123 DYWvinIlKsI-lqLKpGGeLIFIIPdyFP		159 LKkfIinngsFekII
PaeR7I		114 FDlVVGNPPYvRpeIipa	149 DYYipIIsRaLts-LKaGGeLqFIIAdrWM		185 LKslVserIIlkvYV
PstI		146 FNkaIlNPPYlKiaskgR	177 NLYsaFValeI-kqLKsGGeLvaItPrsFc		213 FKrqMldecsIxkIh
TaqI		99 FDlILGNPPYgivgeasK	141 NLYgaFleKaV-rlLKpGGvLvFVVPatWL		177 LKefLaregktsvYY
TthHB8I		97 FDlILGNPPYgivgeasK	139 NLYraFleKaV-rlLKeGGrLvFVVPatWL		175 LKefLaregrtevYY
VspI		211 KDvIPtNPPWrKlpkkd	265 lIqdaGFniasyesVKkAAlanqIVKlldF		310 LKkqcpddqhaticV
EcoRI		133 sDiVVtNPPFsIfreyld	175 NLIksnkiwlgyhlgR--GvsqFIVFehYe		208 aRidengnilIspnn
COMtase		135 lDmVFlDhwk--------	147 yLpdtLLlekcgllrK--GtVlLaDNVlVp		183 VRgsssfecthYssYL
HhaI	(C5)	72 hDiLcagfPCQAfsiagk	99 tLffdI-aRiV-rekK--PkVVFMENVknF		136 VKntMnelDYsFhakV

Protein		Motif VIII	MW	Target	Swissprot Accession Number
AccI	(N6)	209 senVYdal	540 aa	GTMKAC	P25201
BanIII		220 dtkLisav	580 aa	ATCGAT	P22772
BsuBI		223 rdkaIride	501 aa	CTGCAG	P33563
CviBIII		194 dkhdIrdrn	377 aa	TCGA	P10835
Eco57I		202 kelWIadtl	540 aa	CTGAAG	P25240
HincII		177 eskVIsgvs	502 aa	GTYRAC	P17744
PaeR7I		203 dtpeIiadv	531 aa	CTCGAG	P05103
PstI		232 rksaIkasd	507 aa	CTGCAG	P00474
TaqI		192 lgeVIpgkk	421 aa	TCGA	P14385
TthHB8I		190 lgeVIpgrk	428 aa	TCGA	P25749
VspI		333 sqrVIekng	408 aa	ATTAAT	Q03055
EcoRI		228 nidVWirnk	325 aa	GAATTC	P00472
COMtase			221 aa	catechol	P22734
HhaI	(C5)	161 GkReRiymi	327 aa	GCGC	P05102

Figure 1C

linear orders of conserved motifs in their primary sequences. Together with our observation that the AdoMet and methylatable base binding pockets have remarkably similar structures, this suggests catalytic roles for several of the conserved side-chains and has implications for the evolutionary history of these enzymes.

Results

Nine conserved motifs of 5mC Mtases are present in amino Mtases

The structural similarity of the active site and AdoMet-binding regions of M.*Taq*I to those of M.*Hha*I suggests that the amino Mtases contain homologs of the conserved motifs found in 5mC Mtases. The sequences of N6mA and N4mC DNA Mtases were therefore gathered and analyzed as described in Materials and Methods. We were prepared, for two reasons, to look for these motifs in linear orders not seen among the 5mC Mtases. First, others had noted that the two previously identified conserved motifs in amino Mtases appeared in different orders in the various Mtases (Klimasauskas *et al.*, 1989; Wilson & Murray, 1991; Wilson, 1992). Second, others had shown that the 5mC Mtases could function with a circularly permuted motif order (J. Bitinaite, personal communication), or when the regions were expressed separately and allowed to associate *in vivo* (Karreman & de Waard, 1990; Posfai *et al.*, 1991; Lee *et al.*, 1995). We were able to identify nine segments of sequence similarity among the 42 amino Mtases (Figure 1), corresponding to motifs I to VIII and X in the 5mC Mtases (Posfai *et al.*, 1989). We could not identify a homolog to motif IX of the 5mC Mtases; in M.*Hha*I, this motif is involved in the protein folding of the DNA-recognition region (Cheng *et al.*, 1993a).

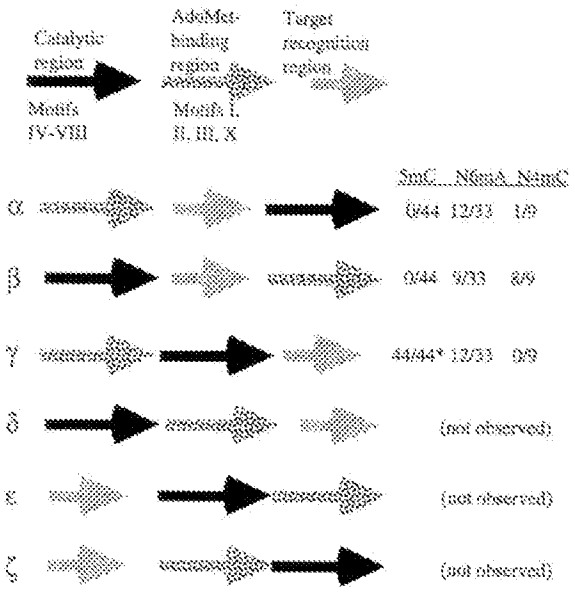
In the structures of M.*Hha*I and M.*Taq*I, motifs I to III and X are primarily responsible for binding AdoMet (Cheng *et al.*, 1993a,b; Klimasauskas *et al.*, 1994; Labahn *et al.*, 1994; Schluckebier *et al.*, 1995), and we term them, collectively, the AdoMet-binding region. The structural comparison suggested that motifs IV, VI, and VIII are primarily responsible for catalysis (Schluckebier *et al.*, 1995), as they form the active site along with motifs V and VII, and we term them collectively the catalytic region.

Three groups of amino Mtases based on motif order

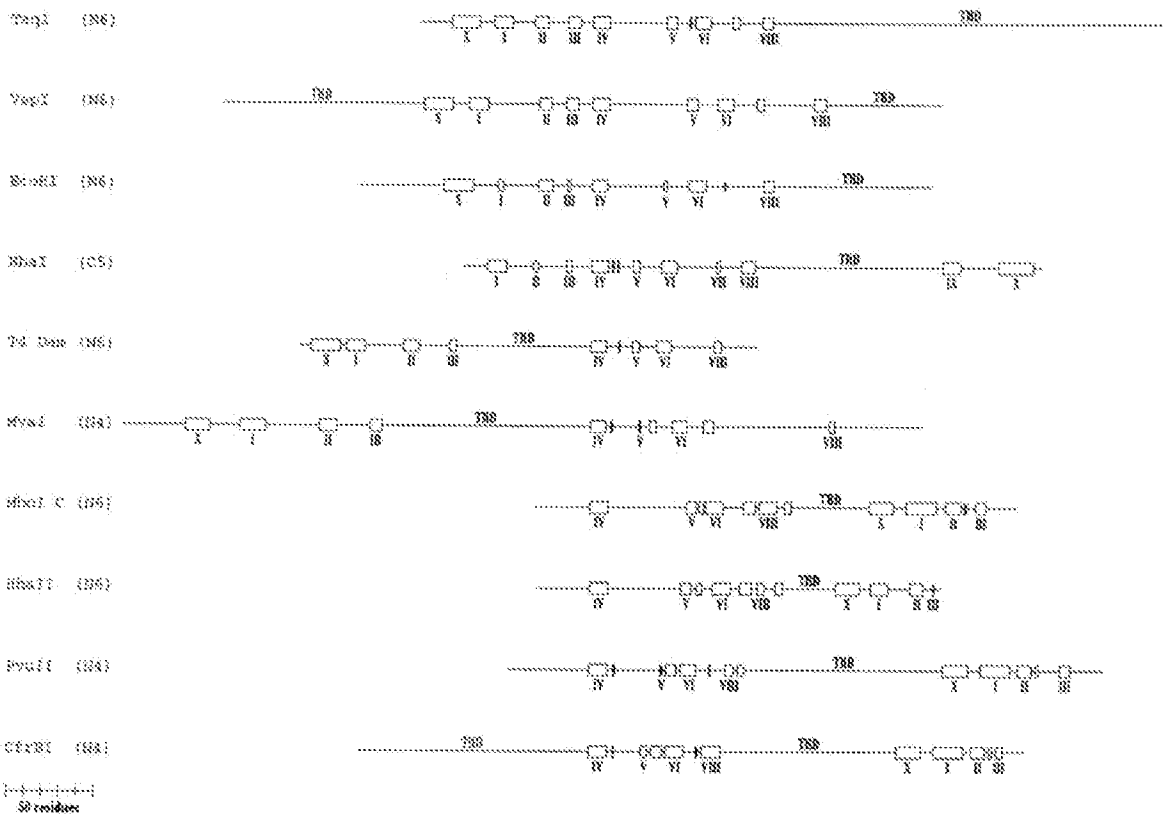
Figure 1 clearly shows that the N6mA Mtases cluster into three distinct groups, based on the order of conserved motifs. It is noteworthy that the N4mC and N6mA Mtases do not group separately from one another. This grouping is compared to earlier analyses in the Discussion.

The validity of the grouping shown in Figure 1, which is based solely on motif order, is supported by the fact that the Mtases within each group are similar to one another by several other criteria as well. First, the groups differ in terms of type of methylation (Figure 2A): N6mA Mtases are found in all three groups, but group β includes eight of the nine N4mC Mtases analyzed, and group γ has a motif order very similar to that seen in the group of all 44 sequenced 5mC Mtases (differing only in the position of motif X; Kumar *et al.*, 1994). Second, comparable motif sequences within each group are more similar to one another than to the same motifs from Mtases in one of the other groups. Third, the groups differ in terms of molecular mass, with group α Mtases being small (260 to 334 amino acid residues), group γ Mtases being large (325 to 580 amino acid residues), and group β covering both of these size ranges (228 to 531 amino acid residues). Fourth, the groups also differ in terms of the DNA sequence recognized. That is, a distinct consensus can be derived for each group: while each group includes specificities that do not match the consensus, and some Mtase specificities could fit more than one consensus, it is clear that the recognition specificities are non-randomly distributed among the groups. As indicated in Figure 1, 12/12 group α N6mA Mtases recognize the sequence (C/G)M$\underline{\text{N}}_{6,8}$T(G/C) (M = A/C; underlining indicates the methylated base), 14/17 group β Mtases recognize the sequence (G/C)N$_{6,8}$M$\underline{\text{N}}_{6,8}$(G/C), and 11/12 group γ Mtases recognize the sequence TNN$\underline{\text{A}}$ (the one exception is M.*Eco*R1). These consensus substrate specificities allow verifiable predictions. For example, the nucleotide sequence of the *Cla*I Mtase gene has not been reported, but its specificity (ATCG$\underline{\text{A}}$T) suggests that it will be found to belong to group γ.

The Mtases differ in the relative linear order of three regions: the AdoMet-binding region, the catalytic (active site) region, and the target recognition region (Figure 2). In the 5mC Mtases, the target recognition region is responsible for specific DNA sequence recognition, and is generally located within the longest gap between conserved motifs (Klimasauskas *et al.*, 1991; Mi & Roberts, 1992; Noyer-Weidner & Trautner, 1993). Group α is arranged in the order (amino to carboxy): AdoMet-binding region, target recognition region, and then catalytic region. Group β is arranged in the order: catalytic region, target recognition region, and AdoMet-binding region. Group γ is arranged in the order: AdoMet-binding region, catalytic region, and target recognition region. No Mtases were found to have the predicted AdoMet binding region between the other two regions (Figure 2A, arrangements δ and ζ). No Mtase is known to have the target recognition region at the amino end (Figure 2A, arrangements ε and ζ); however, M.*Vsp*I (N6mA, assigned to group γ), M.*Cfr*BI and M.*Bam*HI (both N4mC, assigned to group β) have long (>100 amino acid residue) amino-proximal sequences upstream of the first conserved motif, and in theory these upstream sequences could contain the target recognition regions for those three Mtases. If so, M.*Vsp*I could be assigned to group ζ and M.*Cfr*BI and M.*Bam*HI to group ε.



A



B

Figure 2. A, Possible arrangements of the three major regions found in DNA Mtases. To the right, the representation of these arrangements is indicated for Mtases that generate 5mC, N6mA, or N4mC. The asterisk (*) refers to the fact that 5mC Mtases have one major difference from the group γ amino Mtases: motif X is near the carboxy terminus in 5mC Mtases. B, Ten representative examples of 5mC, N6mA, and N4mC Mtases, aligned by motif IV and showing the relative positions of the other conserved motifs. The longest variable region is indicated as the putative target recognition region (labeled as the TRD) in each case.

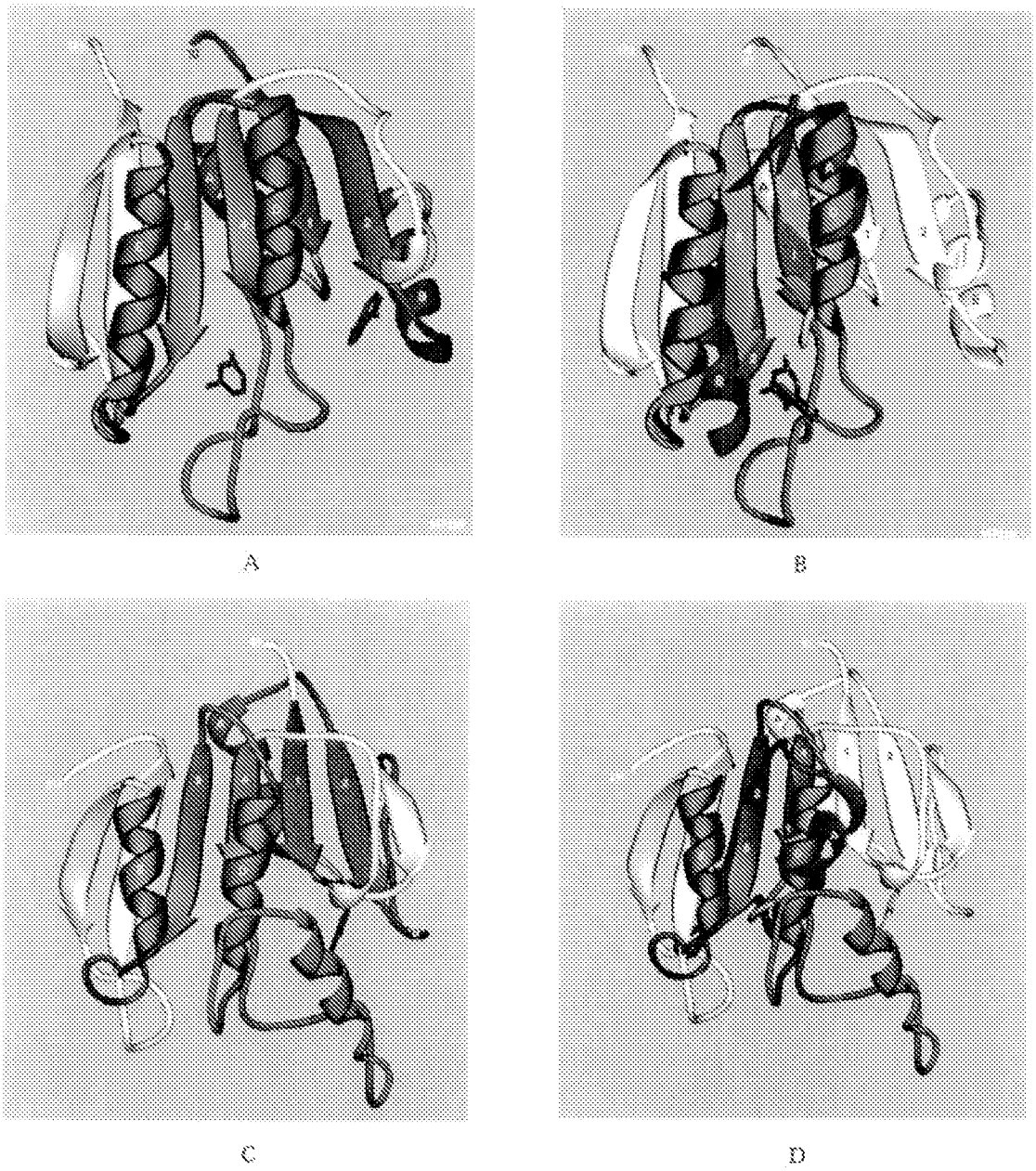


Figure 3. Superimposition of the two α/β clusters from DNA Mtases, in ribbon representation (Carson, 1991). The suggested duplication is not obvious from examination of the primary sequences of M.*Hha*I or M.*Taq*I (but see Laurents *et al.*, 1994). A, B, M.*Hha*I. The two α/β clusters β1 → αA → β2 → αB (shown in green with G-loop in yellow) and β4 → αD → β5 → αE (shown in brown with P-loop in cyan) were isolated from the co-crystal-derived M.*Hha*I–DNA–AdoHcy structure (A), and rotated with respect to one another to achieve the most overlapping possible (B). The β-sheets from the two α/β clusters could be superimposed with an r.m.s.d. of <1 Å for the C^{α} atoms. Also shown are the positions, relative to the respective α/β clusters, of the AdoHcy adenosyl moiety (green) and the target cytosine ring (brown). C, D, M.*Taq*I. Presented as described for M.*Hha*I, except that the target adenine ring is not shown, since its structural position has not yet been determined.

Comparison of conserved motifs among the Mtase families

While the order of conserved motifs varies among the three groups of Mtases, many of their basic features are, by definition, retained. These conserved features are described below.

Motif I

This motif has been described in a variety of AdoMet-dependent Mtases (see Introduction). Structurally, motif I forms the secondary structure β1-loop-αA (Figure 1; Schluckebier *et al.*, 1995). Gly, and less frequently Ala or Pro, form the loop (G-loop) that binds the methionine moiety of

AdoMet. The core of the G-loop is the Gly-X-Gly tripeptide (X is any amino acid), but three group γ Mtases in Figure 1 replace the first Gly with Ala or Ser (M.*Eco*RI), while nine (six of them in group α) replace the second Gly with one of seven alternatives. The majority of amino Mtases have Pro as the last amino acid of strand β1 (Pro46 in M.*Taq*I), but seven of 13 Mtases in group α (and the 5mC Mtases) have Ile or Leu at that position (Leu17 in M.*Hha*I). Conserved hydrophobic side-chains in strand β1 are required for packing against helix αA. The only motif I rule without exceptions among the Mtases in Figure 1 involves the position 4 amino acids upstream of the Gly-X-Gly, which is, in all cases, Asp or Glu. This is the penultimate position of β1 (Figure 1), and poses additional stereochemical constraints by interacting with the dipoles of the peptide bonds in the G-loop.

The most pronounced motif I difference among the Mtases is that both groups α and β, as well as the 5mC Mtases, have Phe at the beginning of the G-loop (second position amino to Gly-X-Gly), but group γ Mtases have Ala, Ser or Gly instead; just two Mtases in Figure 1 violate this pattern (M.*Eco*RI in group γ and M.*Hha*II in group β). In the structure of the 5mC Mtase M.*Hha*I, this G-loop Phe18 forms an edge-to-face van der Waals (vdw) contact with the adenine moiety of AdoMet. However, in the structure of M.*Taq*I (a member of group γ), the same interaction with AdoMet is provided by the Phe146 ring from helix αD (Schluckebier *et al*., 1995). It appears that, in group γ, the Phe that begins the G-loop in the 5mC Mtases is replaced by a spatially equivalent Phe from helix αD/motif V (see below).

Motifs II and III

These two motifs were described as less-conserved blocks in 5mC Mtases (Posfai *et al*., 1989). In the structurally characterized Mtases, motif II contains a negatively charged amino acid at the last position in strand β2, interacting with the ribose hydroxyls of AdoMet, and followed by a bulky hydrophobic side-chain that makes vdw contacts with the AdoMet adenine (Glu40-Trp in M.*Hha*I, Glu71-Ile in M.*Taq*I). Of the 42 amino Mtases in Figure 1, 30 match a (Glu/Asp)-Φ consensus, where Φ is any bulky hydrophobic side-chain, usually followed by Asp, Glu or Asn. The groups do not differ substantially in this.

Motif III also contains an Asp/Glu or Asn/Gln in the first position of αC (Asp80 in M.*Hha*I and Asp89 in M.*Taq*I), which interacts directly with the exocyclic NH_2 (N6) of the AdoMet adenine (Schluckebier *et al*., 1995). Motif III, in addition, provides a hydrogen bond to N1 of the AdoMet adenine from a peptide backbone NH group (Ile81 in M.*Hha*I and Phe90 in M.*Taq*I). The corresponding position is group-specifically conserved in the amino Mtases (Figure 1).

Motif IV

What we call motif IV of the amino Mtases was found in early sequence comparisons and called a "DPPY motif" based on its sequence (Hattman *et al*., 1985; Chandrasegaran & Smith, 1988). A later comparison of 16 N6mA and three N4mC Mtases identified only two conserved segments (Klimasauskas *et al*., 1989), one of which is motif I. The other conserved segment was the DPPY motif which, it was suggested, might correspond to motif IV in 5mC Mtases, even though the reaction mechanisms appear to be quite distinct. The structural comparison of M.*Hha*I and M.*Taq*I have confirmed this correspondence (Schluckebier *et al*., 1995). This diprolyl motif is located in the loop region outside the carboxyl end of β4 (the P-loop; Figure 1). The P-loop forms the active site, along with motifs VI and VIII (see Discussion). The peptide backbone of the corresponding P-loop in M.*Hha*I also contributes to the AdoMet binding site (Cheng, 1995a; see also Kossykh *et al*., 1993). Motif IV has the consensus sequence Asp-Pro-Pro-Tyr (DPPY) in group α, DPPY for N6mA Mtases in group β, Asn-Pro-Pro-Tyr (NPPY) in group γ, and Ser-Pro-Pro-Tyr (SPPY) for N4mC Mtases; this grouping pattern for motif IV has been noted previously (Wilson & Murray, 1991; Wilson, 1992). There are exceptions to this pattern, most notably M.*Bam*HI (N4mC in group β, which has a DPPF), M.*Hha*II (N6mA in group β, DPQY) and M.*Sst*I-a (N6mA in group α, DTPY).

Motif V

In group γ, motif V contains the consensus (Asn/Asp)-Leu-Tyr-X-X-Phe-(Leu/Val/Ile). As described above, in group γ, this Phe replaces the Phe that begins the G-loop in the Mtases of groups α or β. The Leu (Leu100 in M.*Hha*I, Leu142 in M.*Taq*I) makes vdw contacts to the AdoMet adenine on the same side as the Phe (Schluckebier *et al*., 1995). Of the Mtases in Figure 1, only one that lacks Phe at the start of the G-loop fails to contain it in motif V, and that is M.*Hha*II, which has a Ile at that point. In groups α and β, one of the conserved hydrophobic side-chains in motif V may have the same spatial position as the Leu in group γ Mtases.

Motifs VI, VII and VIII

In the structurally characterized Mtases, motif VI forms strand β5 (Figure 1; Schluckebier *et al*., 1995). A conserved Gly starts the strand, and the strand ends with Gly, Pro or Ala, while in group α it ends with Ser-Asn. Groups γ and β also differ from group α in having a conserved pattern of hydrophobic amino acids in β5.

Motif VII is not strongly conserved even among 5mC Mtases, yet credible candidates can be found within each group. In M.*Hha*I, this motif includes Asp144-Tyr in the loop between helix αE and strand β6, and it faces away from the DNA-binding cleft

(Cheng *et al.*, 1993a). It is thus believed to be involved in the folding of the catalytic region (Cheng, 1995b).

In the primary sequence, motif VIII bears little resemblance to the motif present in 5mC Mtases (Gln161-X-Arg-X-Arg165 in M.*Hha*I). This presumably reflects the fact that the 5mC Mtases interact with cytosine *via* hydrogen bonds (through Arg165 in M.*Hha*I), while the N6mA Mtases appear to interact with the target DNA adenine *via* hydrophobic interactions. In the structure of M.*Taq*I, the corresponding region (the loop connecting strands β6 and β7) contains Phe196, which aligns to a conserved Phe or Tyr in other amino Mtases. It is suggested that Phe196 makes favorable edge-to-face or face-to-face vdw contacts to the target DNA adenine (Schluckebier *et al.*, 1995).

Motif X

The location of motif X in the primary sequence is one of the major differences between the 5mC and amino Mtases. In the 5mC Mtases, this motif comes from the carboxy terminus. In the amino Mtases, the corresponding motif is always to the amino side of motif I: at the amino terminus of the protein in groups α and γ, and in the middle of the protein in group β. There are pronounced group-specific differences in the sequence of this motif (Figure 1). However, in all Mtases, this motif is expected to form a helix next to strand β1 (formed by motif I), with conserved hydrophobic side-chains required at certain positions for packing against the β-strands, and a loop preceding the helix (Figure 1). This loop, along with the G-loop of motif I and the P-loop of motif IV, form the sides of the binding pocket in M.*Hha*I for the methionine moiety of AdoMet (Cheng, 1995a).

Discussion

Structural comparison of the Mtase groups

The identification of nine conserved motifs shared with the 5mC Mtases allows the amino Mtases from each group to be mapped onto the consensus structure in a systematic manner (Figure 1). The most pronounced difference among these three groups of amino Mtases is the connection between the proposed AdoMet-binding and catalytic regions. In group γ, a connection between helix αC and strand β4 links the two regions; M.*Taq*I belongs to this group. M.*Hha*I and COMtase also belong to group γ, based on the order of the conserved motifs (Figure 1; excepting motif X in the case of the 5mC Mtase M.*Hha*I), meaning that all currently available Mtase structural information is from Mtases with, essentially, a group γ motif order. In groups α and β, the two regions are apparently connected *via* a separate domain, the target recognition region. The catalytic and AdoMet-binding regions of these Mtases could nevertheless fit the consensus M.*Hha*I–M.*Taq*I structure. Whether this actually occurs is currently being explored, as several more DNA Mtases are undergoing crystallographic analysis. As the group α and β Mtases are proposed to have the DNA recognition domain between the AdoMet-binding and catalytic regions, it is interesting that other structurally characterized proteins have a recognition domain inserted between sequences that form a catalytic β-sheet cluster (for example, the G protein, Coleman *et al.*, 1994; the R.*Pvu*II endonuclease, Cheng *et al.*, 1994), and that two flavoproteins have different domains inserted between parts of the FAD-binding domain (Mittl & Schulz, 1994; Schreuder *et al.*, 1994).

Catalysis of N6-adenine methylation

What can the consensus M.*Hha*I–M.*Taq*I structure and the conservation of nine sequence motifs among the amino and 5mC Mtases tell us about the possible catalytic mechanism of the N6mA Mtases? We propose that the answer to this question lies in a comparison of the binding sites for DNA adenine and for the adenosyl moiety of AdoMet, which are strikingly similar. While no N6mA Mtase–DNA co-crystal structure has yet been determined, the conservation of structure and function among AdoMet-dependent enzymes is supported by both the similar structural framework of the catalytic domains found in M.*Hha*I, M.*Taq*I, and COMtase, and by the similar conformation of the bound AdoMet with the methyl group positioned (not surprisingly) close to the substrate (Schluckebier *et al.*, 1995). This structure-function conservation is also suggested by the conservation of amino acids from motifs I, II, III, V, and X which, in the structurally characterized Mtases, interact with AdoMet.

Are the DNA–adenosyl and AdoMet–adenosyl binding sites structurally comparable? They are each dominated by comparable α/β clusters (β1 → αA → β2 → αB and β4 → αD → β5 → αE); the former includes motifs I and II, and forms the bulk of the AdoMet-binding region, and the latter includes motifs IV to VI, and forms the bulk of the catalytic region. These two α/β clusters and their bound substrates do, in fact, have strikingly similar structures. The two α/β clusters from the M.*Hha*I–DNA–*S*-adenosyl-L-homocysteine (AdoHcy) structure can be superimposed, with a root-mean-square deviation (r.m.s.d.) of <1 Å for the C^{α} atoms in the β strands, with the AdoHcy adenosyl moiety overlapping the target cytosine ring (Figure 3A and B). Similar overlapping is also possible for the α/β clusters of the M.*Taq*I–AdoMet structure (Figure 3C and D) and of the M.*Hae*III–DNA structure (results not shown). While the various Mtase groups have these two α/β clusters in different orders (Figures 1 and 2), in no case has the motif order rearrangement interrupted an α/β cluster (Figure 1). The relatedness of the binding pockets for the DNA base and for AdoMet may

explain an interesting feature of the M.*Hae*III structure (Reinisch *et al.*, 1995): the unpaired 5' thymine of one DNA duplex penetrates the AdoMet pocket of the neighboring Mtase–DNA complex. The thymine does not enter deeply enough to interact with the conserved acidic amino acid side-chains (see section on motifs II and III), but does make the hydrophobic contacts made by the AdoMet adenosyl moiety, such as face-to-face stacking with Tyr30 of motif II (analogous to Trp41 in M.*Hha*I and Ile72 in M.*Taq*I).

Based on the chemical and structural similarity of the DNA–adenosyl and AdoMet–adenosyl moieties, and the structural similarity of the AdoMet-binding and catalytic regions of the Mtase, we propose analogous Mtase–adenosine interactions in the two regions (Table 1). The methylation of adenine appears to result from a direct attack of the AdoMet methyl group on the adenine N6 (Pogolotti *et al.*, 1988; Ho *et al.*, 1991).

In analogy to the hydrogen bond between AdoMet–adenosyl N6 and a motif III Asp in M.*Hha*I and M.*Taq*I, we suggest that the N6 amino nitrogen of the target adenine is the donor in a hydrogen bond to the side-chain of Asp/Asn in motif IV, and possibly to one of the main-chain oxygens of the adjacent two proline residues. This would negatively polarize N6, activating it for direct transfer of the CH_3 from AdoMet. In Mtases with Asn in this position (group γ) the carboxamide could be the donor in a hydrogen bond to adenine N1, as well as an acceptor from adenine N6, similar to the role Asn229 of thymidylate synthase plays in hydrogen bonding to dUMP (Liu & Santi, 1993; also see Figure 3 of Gerlt, 1994). Mtases with Asp in this position could also hydrogen bond adenine N1 if the carboxyl is protonated.

Consistent with the above, mutation of DPPY to GPPY or APPY in the two halves of the bifunctional Mtase M.*Fok*I (group α) abolishes activity in the altered half (Sugisaki *et al.*, 1989). Altering DPPY to SPPY or NPPY abolishes the activity of the group α N6mA Mtase M.*Eco*Dam (Guyot *et al.*, 1993). The M.*Eco*Dam result is somewhat surprising, as NPPY is in motif IV in nine of 12 Mtases of group γ (Figure 1), while SPPY is in motif IV in six Mtases of group β and even one (M.*Mva*I) of group α (see the following section on N4mC Mtases). Similarly, altering NPPF to DPPF in M.*Eco*KI led to loss of activity (Willcock *et al.*, 1994). (M.*Eco*KI is a type I N6mA Mtase that has also been modeled onto the consensus M.*Hha*I–M.*Taq*I structure (Dryden, *et al.*, 1995) We interpret these data from mutant enzymes to mean that the relative positions of the activating hydrogen bond acceptor, target amino group, and AdoMet methyl group must be precisely maintained.

The two proline residues are not present in all examples of motif IV: M.*Hha*II (group β; DPQY) and M.*Sts*I-a (group α; DTPY) resemble the rRNA N6mA Mtases in this respect. Analysis of 12 rRNA N6mA Mtases (EC 2.1.1.48) reveals the consensus (N/S)PP(Y/F) (X. Cheng, unpublished observations). Altering motif IV of the bacteriophage T4 Dam N6mA Mtase from DPPY to DAPY or DTPY (as occurs in M.*Sts*I-a) substantially increased K_D^{AdoMet}, but had much smaller effects on k_{cat}. K_D^{AdoMet} and K_D^{DNA} (Kossykh *et al.*, 1993). This suggests that the Pro alteration affects a catalytic but not a rate-limiting (not k_{cat}-determining) step, consistent with the above inferences (particularly if product release is the rate-limiting step, as it is for at least some Mtases: Reich & Mashoon, 1993). Unfortunately, these mutations have not been made in M.*Eco*Dam, but in that enzyme, changing DPPY to DGPY, DVPY, DPGY, DPRY, DPQY (as occurs in M.*Hha*II), DPEY, or DPVY all abolished activity (Guyot *et al.*, 1993). In summary, these results are

Table 1. Comparison of the DNA–adenosyl and AdoMet–adenosyl binding sites in group γ[a]

	Target adenine ring			AdoMet adenine ring			Possible functions
	Motif	aa	Location	Motif	aa	Location	
(A)	IV	Asn	First aa of P-loop	III	Asp/Ser	First aa of αC	Side-chain hydrogen bonds to N6-nitrogen.[b]
(B)	"	"	"	III	Phe/Tyr	Second aa of αC	Main-chain NH group hydrogen bonds to N1-nitrogen.
(C)	IV	Tyr/Phe/Trp	Forth aa of P-loop	V[d]	Phe[d]	αD[d]	Edge-to-face vdw contact with adenine.
(D)	VI	Ile/Val	Penultimate aa of β5	V	Leu/Ile/Tyr	Last aa of P-loop	vdw contact with adenine ring on the same face as to (C).
(E)	VIII	Phe	Loop between strands β6 and β7	II	Ile/Val/Leu/Phe	First aa of loop between strand β2 and helix αB	Face-to-face vdw contact with adenine ring on the opposite face.

[a] See also Figure 1C.

[b] N6 of target adenine could form a second hydrogen bond to one of the main-chain oxygen atoms of the two proline residues in motif IV at the P-loop.

[c] The amide side-chain of Asn in (A) could be a hydrogen bond donor to N1 of adenine. If the carboxyl of Asp is protonated, it could also be hydrogen bond donor to adenine N1 (N6mA Mtases in groups α and β) or to cytosine N3 (N4mC Mtase M. *Bam*HI). For N4mC Mtases with Ser at the first position of the P-loop, a conserved Asn from motif VI (in the end of strand β5) could possibly hydrogen bond cytosine N3 in analogy to Glu119 in motif VI of M.*Hha*I.

[d] In groups α and β, the phenyl ring is from motif I, the first amino acid (aa) of the G-loop.

consistent with the role suggested for motif IV, but also make clear the dangers of considering the motifs in isolation.

The Tyr in motif IV, Phe in motif VIII, and hydrophobic side-chains in motif VI could function in properly orienting the target DNA adenine. The analogy for this Tyr:Phe pair is the Phe from motif V (group γ) or motif I (groups α and β) that makes an edge-to-face vdw contact with the AdoMet adenine (Table 1), and to the Trp41 of motif II in M.*Hha*I, which makes a face-to-face vdw contact with the AdoMet adenine. This interpretation is consistent with the fact that altering motif IV of the M.*Eco*KI N6mA Mtase from NPPF to NPPG or NPPC abolished activity without greatly affecting the affinity for DNA or AdoMet (Willcock *et al.*, 1994). In contrast, altering the NPPF to NPPY or NPPW (as occurs in M.*Vsp*I) resulted in an enzyme that retained partial activity (Willcock *et al.*, 1994).

Catalysis of N4-cytosine methylation

From the first report of an N4mC Mtase sequence, it has been suggested, based on overall sequence similarity and the very similar chemical properties of adenine N6 and cytosine N4 (e.g. see Figure 3 of Wetner *et al.*, 1984; Renugopalakrishnan *et al.*, 1971), that N4mC and N6mA Mtases may use a common reaction mechanism (Tao *et al.*, 1989). We believe that N4-cytosine and N6-adenine methylation do use the same catalytic mechanism, for the following reasons. First, as described above, the N6mA and N4mC Mtases in group β appear to be more closely related to one another than either subgroup is to the N6mA Mtases of group γ. Second, the N6mA and N4mC versions of motifs are either indistinguishable from one another, or are not consistently different from one another (the exceptions are usually provided by M.*Mva*I or M.*Bam*HI). For example, the most obvious difference between the N4mC and N6mA Mtase sequences is the conserved Ser present in motif IV in place of Asp/Asn; yet this Ser must not represent an essential functional difference, as it is not present in the N4mC Mtase M.*Bam*HI. We suggest that the Ser in motif IV of most N4mC Mtases could hydrogen bond and activate the amino N4 nitrogen of the cytosine ring, in analogy to COMtase, in which a Ser is hydrogen-bonded to the AdoMet N6 (Vidgren *et al.*, 1994), and in analogy to the Asp in motif III of the N6mA Mtases (Table 1). These Mtases may also be hydrogen bond donors to cytosine N3, similar to the proposed bonding of N6mA Mtases to adenine N1, but the donor side-chain would be from outside motif IV (possibly a conserved Asn in motif VI). Other requirements for N4 methylation of cytosine may be implied by the fact that the individual motifs in M.*Mva*I, while unambiguously in the group α order, more closely resemble the corresponding motifs of the other (group β) N4mC Mtases (Figure 1).

DNA Mtase families and comparison to earlier Mtase groupings

It should be noted that the results of our analysis are very consistent with, and provide a structural basis for, earlier attempts to group Mtases. The first of these attempts examined 17 Mtases, and was based not on motif identification and order but on overall sequence alignment (Chandrasegaran & Smith, 1988). That analysis found five groups of Mtases. Their group I included four Mtases all in our group α; their group II included two Mtases both in group β; their group III included four Mtases all in group γ; their group IV included six 5mC MTases (which we did not include, but which have a group γ motif order, except for the position of motif X); and their group V consisted of M.*Eco*RI, which has several variations from the consensus motifs, but which we assign to group γ based on motif order. The second major analysis categorized 33 type II amino Mtases (Wilson & Murray, 1991). They placed the amino Mtases into five groups, based on the order and nature of just motifs I and IV (again, M.*Eco*RI was not grouped). Their analysis is consistent with our assignments except that, (1) they grouped the N4mC and N6mA Mtases separately, and (2) they assigned M.*Sma*I differently. The N4mC and M6mA Mtases were not separated in a later version of that analysis (Wilson, 1992), and we have adopted the nomenclature of that later analysis. Four to ten Mtases from group γ were also clustered by Lauster *et al.* (1987), Janulaitis *et al.* (1992), and Noyer-Weidner *et al.* (1994), based on overall sequence similarity. More recently, Timinskas *et al.* (1995), using a sensitive method to make pairwise comparisons between amino Mtases, detected two conserved motifs in addition to the two that had been identified (see Klimasauskas *et al.*, 1989). Their conserved motif (CM) Is corresponds to motif X, CM I to motif I, CM II to motif IV, and CM III to motifs V and VI. Using these, they have defined eight groupings for the Mtases which are consistent with ours and with Wilson & Murray (1991); Timinskas *et al.* (1995) have just subdivided their groupings to separate N4mC and N6mA Mtases. Since our analysis began with assignment of motifs I and IV (Materials and Methods), it is not surprising that we got our grouping consistent with Wilson & Murray (1991), but the discovery of the seven additional conserved motifs in consistent orders greatly strengthens the basis for this Mtase categorization.

Evolutionary implications

The four DNA Mtase arrangements seen to date (α, β, γ, and 5mC) differ in the linear order of conserved motifs (Figure 2), but in no case are either of the two α/β clusters interrupted. Furthermore, consider the superimposable structures of those clusters, which appear to have the same linear order in all four groups (β1/β4 → αA/αD → β2/β5 → αB/αE; Figure 3), and the apparent functional

relatedness of the AdoMet and adenine binding sites (Table 1). All of this is consistent with the possibility that the original Mtases arose after gene duplication converted an AdoMet-binding protein into a protein that bound two molecules of AdoMet (see also Lauster, 1988, 1989; Tao *et al.*, 1989; Cuyot & Caudron, 1994). The two halves could have diverged when either the amino-proximal or carboxy-proximal AdoMet-binding site evolved to bind adenine, and then diverged further to yield COMtase (Figure 1) and the wide variety of other AdoMet-dependent Mtases (Fujioka, 1992; Clarke, 1993). To become DNA Mtases, by this duplication model, an additional fusion would have brought the target recognition region in at the carboxy terminus (group γ) or between the AdoMet and adenine binding sites (groups α and β) of the ancestral adenine Mtase. This duplication model may provide insight into the fact that some Mtases appear to bind two molecules of AdoMet (Bergerat & Guschlbauer, 1990; Adams & Blumenthal, 1995), one of which affects the selectivity between substrate and non-specific DNA sequences (for M.*EcoDam*; Bergerat & Guschlbauer, 1990). It is also noteworthy, with regard to the proposed importation of a target recognition region, that some 5mC Mtases are naturally made as two separate polypeptides: one has motifs I to VIII (including both α/β clusters) and the other carries the target-recognizing region and motifs IX and X, and these associate in the cell to form active enzyme (Karreman & de Waard, 1990; Lee *et al.*, 1995).

While the 5mC and N4mC Mtases each fall into single groups defined by motif order (with one exception, M.*Mva*I: N4mC Mtase assigned to group α), the N6mA Mtases we examined are fairly evenly distributed among the three groups (27% in group β, and just over 36% each in groups α and γ). This may also be explained by a model in which the original nucleic acid Mtase(s) generated N6mA.

In summary, the DNA Mtases appear to be, paradoxically, both more uniform (shared conserved motifs; N4mC Mtases not a distinct group) and more diverse (four possible motif orders) than had been expected. The solved structures of Mtases from groups α and β should be very informative.

Materials and Methods

The Swissprot database was searched by EC number, yielding the amino acid sequences of 33 N6mA DNA Mtases (EC 2.1.1.72) and nine N4mC DNA Mtases (EC 2.1.1.113). The names and accession numbers of these Mtases are listed in *Figure 1*. For comparison, we also used the sequences of the 5mC Mtase (EC 2.1.1.73) M.*Hha*I and the small-molecule COMtase (EC 2.1.1.6).

A scan of the sequences was first performed to locate motif I and motif IV. These two blocks provided the anchor points for global alignment of other motifs. As comparable motifs did not always appear in the same linear order, the alignments were refined within each Mtase group. Our analysis of M.*Eco*RI yielded a different motif I assignment from that of Klimasauskas *et al.* (1989). For clarity and convenience, we retain the nomenclature of Posfai *et al.* (1989) for the 5mC Mtase conserved motifs and of Wilson (1992) for the Mtase groups. The M.*Hha*I–M.*Taq*I structural alignment was crucial to this analysis, as it indicated which motif positions were most functionally significant and which substitutions were likely to be permissible.

Acknowledgements

We thank Catherine Luschinsky Drennan (University of Michigan) for critical discussion of the structural implications, Markus Wilmanns and David T. F. Dryden (University of Edinburgh) for discussion of the group assignments, Gerd Schluckebier and Wolfram Saenger (Freie Universitat) for the coordinates of M.*Taq*I, and Karin M. Reinisch and William N. Lipscomb (Harvard University) for the coordinates of M.*Hae*III. We also thank Ashok S. Bhagwat and Joan C. Dunbar (Wayne State University), David T. F. Dryden (University of Edinburgh), Stanley Hattman (University of Rochester), Rowena G. Matthews and Joseph T. Jarrett (University of Michigan), and Richard J. Roberts, Sanjay Kumar, Janos Posfai, and Geoffrey G. Wilson (New England Biolabs) for helpful discussions and comments on the manuscript. This work was supported by the W. M. Keck Foundation (to X.C.), by grant GM49245 from the National Institutes of Health (to X.C.), and by grant DMB-9205248 from the National Science Foundation (to R.B.).

References

Adams, C. & Blumenthal, R. M. (1995). What is the significance of the internal duplication in the *Pvu*II DNA methyltransferase? *FASEB J.* **9**, A1389.

Bergerat, A. & Guschlbauer, W. (1990). The double role of methyl donor and allosteric effector of *S*-adenosylmethionine for Dam methylase of *E. coli*. *Nucl. Acids Res.* **18**, 4369–4375.

Bestor, T. H., Laudano, A., Mattaliano, R. & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. *J. Mol. Biol.* **203**, 971–983.

Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallog.* **24**, 958–961.

Chandrasegaran, S. & Smith, H. O. (1988). Amino acid sequence homologies among twenty-five restriction endonucleases and methylases. In *Structure & Expression: From Proteins to Ribosomes* (Sarma, R. H. & Sarma, M. H., eds), vol. 1, pp. 149–156, Adenine Press, New York.

Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S. & Verdine, G. L. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry*, **30**, 11018–11025.

Chen, L., MacMillan, A. M. & Verdine, G. L. (1993). Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase. *J. Am. Chem. Soc.* **115**, 5318–5319.

Cheng, X. (1995a). DNA modification by methyltransferases. *Curr. Opin. Struct. Biol.* **5**, 4–10.

Cheng, X. (1995b). Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293–308.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. (1993a). Crystal structure of the *Hha*I methylase

complexed with S-adenosyl-L-methionine. *Cell*, **74**, 299–307.

Cheng, X., Kumar, S., Klimasauskas, S. & Roberts, R. J. (1993b). Crystal structure of the *Hha*I DNA methyltransferase. *Cold Spring Harbor Symp. Quant. Biol.* **58**, 331–338.

Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J. E. (1994). Structure of *Pvu*II endonuclease with cognate DNA. *EMBO J.* **13**, 3927–3935.

Clarke, S. (1993). Protein methylation. *Curr. Opin. Cell. Biol.* **5**, 977–983.

Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G. & Sprang, S. R. (1994). Structures of active conformations of Gi*α*1 and the mechanism of GTP hydrolysis. *Science*, **265**, 1405–1412.

Dryden, D. T. F., Murrock, S. S. & Winter, M. (1995). Structural modelling of a type I DNA methyltransferase. *Nature Struct. Biol.* **2**, 632–635.

Finnegan, E. J. & Dennis, E. S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucl. Acids Res.* **21**, 2383–2388.

Friedman, S. & Ansari, N. (1992). Binding of the *Eco*RII methyltransferase to 5-fluorocytosine-containing DNA. Isolation of a bound protein. *Nucl. Acids Res.* **20**, 3241–3248.

Fujioka, M. (1992). Mammalian small molecule methyltransferases: their structural and functional features. *Int. J. Biochem.* **24**, 1917–1924.

Gerit, J. A. (1994). Protein engineering to study enzyme catalytic mechanisms. *Curr. Opin. Struct. Biol.* **4**, 593–600.

Guenthner, P., Kim, S. & Badmarik, D. P. (1992). Cloning, isolation, and characterization of the human T-cell DNA-cytosine 5-methyltransferase gene. *J. Cell. Biochem. Suppl.* **1992**, 179.

Guyot, J. B. & Caudron, B. (1994). Statistical significance of amino acid sequence similarity in type II DNA methyltransferases. *C.R. Acad. Sci. III, Sci. Vie.* **317**, 20–24.

Guyot, J.-B., Grassi, J., Hahn, U. & Guschlbauer, W. (1993). The role of the preserved sequences of Dam methylase. *Nucl. Acids Res.* **21**, 3183–3190.

Hanck, T., Schmidt, S. & Pritz, H. J. (1993). Sequence-specific and mechanism-based crosslinking of Dam DNA cytosine-C5 methyltransferase of *E. coli* K-12 to synthetic oligonucleotides containing 5-fluoro-2′-deoxycytidine. *Nucl. Acids Res.* **21**, 303–309.

Hartman, S., Wilkinson, J., Swinton, D., Schlagman, S., Macdonald, P. M. & Mosig, G. (1985). Common evolutionary origin of the phage T4 dam and host *Escherichia coli* dam DNA-adenine methyltransferase genes. *J. Bacteriol.* **164**, 932–937.

Ho, D. K., Wu, J. C., Santi, D. V. & Floss, H. G. (1991). Stereochemical studies of the C-methylation of deoxycytidine catalyzed by *Hha*I methylase and the N-methylation of deoxyadenosine catalyzed by *Eco*RI methylase. *Arch. Biochem. Biophys.* **284**, 264–269.

Ingrosso, D., Fowler, A. V., Bleibaum, J. & Clarke, S. (1989). Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes, common sequence motifs for protein, DNA, RNA, and small molecule *S*-adenosylmethionine-dependent methyltransferases. *J. Biol. Chem.* **264**, 20131–20139.

Janulaitis, A., Vaisvila, R., Timinskas, A., Klimasauskas, S. & Butkus, V. (1992). Cloning and sequence analysis of the genes coding for *Eco*571 type IV restriction-modification enzymes. *Nucl. Acids Res.* **20**, 6051–6056.

Kagan, R. M. & Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse *S*-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* **310**, 417–427.

Karreman, C. & de Waard, A. (1990). *Agmenellum quadruplicatum* M.*Aqu*I, a novel modification methylase. *J. Bacteriol.* **172**, 266–272.

Kita, K., Kotani, H., Ohta, H., Yanase, H. & Kato, N. (1992). *Sta*I, a new *Fok*I isoschizomer from *Streptococcus sanguis* 54, cleaves 5′ CGATC (N) 10/14 3′. *Nucl. Acids Res.* **20**, 9999.

Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. & Janulaitis, A. (1989). Sequence motifs characteristic of DNA (cytosine-N4) methyltransferases, similarity to adenine and cytosine-C5 DNA-methylases. *Nucl. Acids Res.* **17**, 9823–9832.

Klimasauskas, S., Nelson, J. L. & Roberts, R. J. (1991). The sequence specificity domain of cytosine-C5 methylases. *Nucl. Acids Res.* **19**, 6183–6190.

Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.

Kossykh, V. G., Schlagman, S. L. & Hattman, S. (1993). Conserved sequence motif DPPY in region IV of the phage T4 Dam DNA-[N6-adenine]-methyltransferase is important for *S*-adenosyl-L-methionine binding. *Nucl. Acids Res.* **21**, 3563–3566.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J. & Wilson, G. G. (1994). The DNA (cytosine-5) methyltransferases. *Nucl. Acids Res.* **22**, 1–10.

Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I. & Saenger, W. (1994). Three-dimensional structure of the adenine specific DNA methyltransferase M.*Taq*I in complex with the cofactor *S*-adenosylmethionine. *Proc. Natl Acad. Sci. USA*, **91**, 10957–10961.

Laurents, D. V., Subbiah, S. & Levitt, M. (1994). Different protein sequences can give rise to highly similar folds through different stabilizing interactions. *Protein Sci.* **3**, 1938–1944.

Lauster, R. (1988). Duplication and variation as a phylogenetic principle of type II DNA methyltransferases. *Gene*, **74**, 243.

Lauster, R. (1989). Evolution of type II DNA methyltransferases: a gene duplication model. *J. Mol. Biol.* **206**, 313–321.

Lauster, R., Kriebardis, A. & Guschlbauer, W. (1987). The GATATC-modification enzyme *Eco*RV is closely related to the GATC-recognizing methyltransferases *Dpn*II and dam from *E. coli* and phage T4. *FEBS Letters*, **220**, 167–170.

Lauster, R., Trautner, T. A. & Noyer-Weidner, M. (1989). Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* **206**, 305–312.

Lee, K.-F., Kam, K.-M. & Shaw, P.-C. (1995). A bacterial methyltransferase M.*Eco*HK311 requires two proteins for *in vitro* methylation. *Nucl. Acids Res.* **23**, 103–109.

Liu, L. & Santi, D. V. (1993). Asparagine 229 in thymidylate synthase contributes to, but is not essential for, catalysis. *Proc. Natl Acad. Sci. USA*, **90**, 8604–8608.

Mi, S. & Roberts, R. J. (1992). How M.*Msp*I and M.*Hpa*II decide which base to methylate. *Nucl. Acids Res.* **20**, 4811–4816.

Mi, S. & Roberts, R. J. (1993). The DNA binding affinity of *HhaI* methylase is increased by a single amino acid substitution in the catalytic center. *Nucl. Acids Res.* **21**, 2459–2464.

Mittl, P. R. & Schulz, G. E. (1994). Structure of glutathione reductase from *Escherichia coli* at 1.66 Å resolution: comparison with the enzyme from human erythrocytes. *Protein Sci.* **3**, 799–809.

Noyer-Weidner, M. & Trautner, T. A. (1993). Methylation of DNA in prokaryotes. In *DNA Methylation. Molecular Biology and Biological Significance* (Jost, J. P. & Saluz, H. P., eds), pp. 39–108. Birkhauser Verlag, Basel, Switzerland.

Noyer-Weidner, M., Walter, J., Terschueren, P., Chai, S. & Trautner, T. A. (1994). M.Θ3TII, a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family of adenine-N6-DNA-methyltransferases. *Nucl. Acids Res.* **22**, 4066–4072.

O'Gara, M., McCloy, K., Malone, T. & Cheng, X. (1995). Structure-based sequence alignment of three AdoMet-dependent methyltransferases. *Gene*, **157**, 135–138.

Pogolotti, A. L., Ono, A., Subramantam, R. & Santi, D. V. (1988). On the mechanism of DNA-adenine methylase. *J. Biol. Chem.* **263**, 7461–7464.

Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucl. Acids Res.* **17**, 2421–2435.

Posfai, G., Kim, S. C., Szilak, L., Kovacs, A. & Venetianer, P. (1991). Complementation by detached parts of GGCC-specific DNA methyltransferases. *Nucl. Acids Res.* **19**, 4843–4847.

Reich, N. O. & Mashoon, N. (1993). Pre-steady state kinetics of an *S*-adenosylmethionine-dependent enzyme. *J. Biol. Chem.* **268**, 9191–9193.

Reich, N. O., Maegley, K. A., Shoemaker, D. D. & Everett, E. (1991). Structural and functional analysis of *EcoRI* DNA methyltransferase by proteolysis. *Biochemistry*, **30**, 2940–2946.

Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. (1995). The crystal structure of *HaeIII* methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.

Renugopalakrishnan, V., Lakshminarayanan, A. V. & Sasisekharan, V. (1971). Stereochemistry of nucleic acids and polynucleotides. III. Electronic charge distribution. *Biopolymers*, **10**, 1159–1167.

Santi, D. V., Garrett, C. E. & Barr, P. J. (1983). On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell*, **33**, 9–10.

Santi, D. V., Norment, A. & Garrett, C. E. (1984). Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc. Natl Acad. Sci. USA*, **81**, 6993–6997.

Scheidt, G., Graessmann, A. & Weber, H. (1991). Cloning of the methyltransferase gene from *Arabidopsis thaliana*. *Biol. Chem. Hoppe-Seyler*, **372**, 742.

Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16–20.

Schreuder, H. A., Mattevi, A., Obmolova, G., Kalk, K. H., Hol, W. G., van der Bolt, F. J. & van Berkel, W. J. (1994). Crystal structures of wild-type *p*-hydroxybenzoate hydroxylase complexed with 4-aminobenzoate,2,4-dihydroxybenzoate, and 2-hydroxy-4-aminobenzoate, and of the Tyr222Ala mutant complexed with 2-hydroxy-4-aminobenzoate. Evidence for a proton channel and a new binding mode of the flavin ring. *Biochemistry*, **33**, 10161–10170.

Smith, H. O., Annau, T. M. & Chandrasegaran, S. (1990). Finding sequence motifs in groups of functionally related proteins. *Proc. Natl Acad. Sci. USA*, **87**, 826–830.

Smith, S. S., Kaplan, B. E., Sowers, L. C. & Newman, E. M. (1992). Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proc. Natl Acad. Sci. USA*, **89**, 4744–4748.

Som, S., Bhagwat, A. S. & Friedman, S. (1987). Nucleotide sequence and expression of the gene encoding the *EcoRII* modification enzyme. *Nucl. Acids Res.* **15**, 313–332.

Sugisaki, H., Kita, K. & Takanami, M. (1989). The *FokI* restriction-modification system. II. Presence of two domains in *FokI* methylase responsible for modification of different DNA strands. *J. Biol. Chem.* **264**, 5757–5761.

Tao, T., Walter, J., Brennan, K. J., Cotterman, M. M. & Blumenthal, R. M. (1989). Sequence, internal homology and high-level expression of the gene for a DNA-(cytosine N4)-methyltransferase, M.Pvu II. *Nucl. Acids Res.* **17**, 4161–4175.

Timinskas A., Butkus, V. & Janulaitis, A. (1995). Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene*, **157**, 3–11.

Vidgren, J., Svensson, L. A. & Liljas, A. (1994). Crystal structure of catechol *O*-methyltransferase. *Nature*, **368**, 354–358.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.

Willcock, D. F., Dryden, D. T. F. & Murray, N. E. (1994). A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* **13**, 3902–3908.

Wilson, G. G. (1992). Amino acid sequence arrangements of DNA-methyltransferases. *Methods Enzymol.* **216**, 259–279.

Wilson, G. G. & Murray, N. E. (1991). Restriction and modification systems. *Annu. Rev. Genet.* **25**, 585–627.

Wu, J. C. & Santi, D. V. (1987). Kinetic and catalytic mechanism of *HhaI* methyltransferase. *J. Biol. Chem.* **262**, 4778–4786.

Wyszynski, M. W., Gabbara, S. & Bhagwat, A. S. (1992). Substitutions of a cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. *Nucl. Acids Res.* **20**, 319–326.

Edited by D. E. Draper

(Received 14 June 1995; accepted 11 August 1995)